SOURCE AND ACCURACY STATEMENT FOR THE 1991 WAVE 6+ PUBLIC USE FILES FROM THE SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP)

SOURCE OF DATA

The SIPP universe is the noninstitutionalized resident population living in the United States. The population includes persons living in group quarters, such as dormitories, rooming houses, and religious group dwellings. Not eligible to be in the survey are crew members of merchant vessels, Armed Forces personnel living in military barracks, and institutionalized persons, such as correctional facility inmates and nursing home residents. Also, not eligible are United States citizens residing abroad. Foreign visitors who work or attend school in this country and their families are eligible; all others are not eligible. With the exceptions noted above, field representatives interview eligible persons who are at least 15 years of age at the time of the interview.

The 1991 panel of the SIPP sample is located in 230 Primary Sampling Units (PSUs) each consisting of a county or a group of contiguous counties. Within these PSUs, we systematically selected expected clusters of two living quarters (LQs) from lists of addresses prepared for the 1980 decennial census to form the bulk of the sample. To account for LQs built within each of the sample areas after the 1980 census we selected a sample containing clusters of four LQs from permits issued for construction of residential LQs up until shortly before the beginning of the panel.

In jurisdictions that have incomplete addresses or don't issue building permits, we sampled small land areas, listed expected clusters of four LQs, and then subsampled. In addition, we selected a sample of LQs from a supplemental frame that included LQs identified as missed in the 1980 census.

Approximately 19,300 living quarters were originally designated for the 1991 panel. For Wave 1 of the panel, we obtained interviews from occupants of about 14,300 of the 19,300 designated living quarters. We found most of the remaining 5,000 living quarters in the panel to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. However, we did not interview approximately 1,300 of the 5,000 living quarters in the panel because the occupants refused to be interviewed, could not be found at home, were temporarily absent, or were otherwise unavailable. Thus, occupants of about 92 percent of all eligible living quarters participated in the first interview of the panel.

For subsequent interviews, only original sample persons (those in Wave 1 sample households and interviewed in Wave 1) and persons living with them are eligible to be interviewed. We followed original sample persons if they moved to a new address, unless the new address was more than 100 miles from a SIPP sample area, we attempted telephone interviews. When original sample persons moved to remote parts of the country and were unreachable by telephone, moved without leaving a forwarding address, or refused the interview, additional noninterviews resulted.

The Bureau divides sample households within a given panel into four subsamples of nearly equal size. We call these subsamples rotation groups 1, 2, 3, or 4 and interview one rotation group each month. Beginning in February 1991, we schedule interviews for each household in the sample at 4 month intervals over a period of roughly 2½ years. The reference period for the questions is the 4-month period preceding the interview month. A wave is one cycle of four interviews covering the entire sample, using the same questionnaire.

A unique feature of the SIPP design is overlapping panels. The overlapping design allows combining of panels and essentially doubles the sample size. It is possible to combine selected interviews for the 1991 panels with interviews from the 1990 panels. We include information necessary to do this later in this statement.

The public use files include core and supplemental (topical module) data. Field representatives repeat core questions at each interview over the life of the panel. Topical modules include questions which are asked only in certain waves. The 1991 and 1990 panel topical modules are shown in tables 1 and 2 respectively.

Tables 3 and 4 indicate the reference months and interview months for the collection of data from each rotation group for the 1991 and 1990 panels respectively. For example, Wave 1 rotation group 2 of the 1991 panel was interviewed in February 1991 and data for the reference months October 1990 through January 1991 were collected.

Estimation. We derived SIPP person weights in each panel from several stages of weight adjustments. In the first wave, we gave each person a base weight equal to the inverse of his/her probability of selection. For each subsequent interview, the Bureau gave each person a base weight that accounted for following movers.

We applied a factor to each interviewed person's weight to account for the SIPP sample areas not having the same population distribution as the strata they are from.

We applied a noninterview adjustment factor to the weight of every occupant of interviewed households to account for persons in noninterviewed occupied households which were eligible for the sample. (The Bureau treated individual nonresponse within partially interviewed households with imputation. We made no special adjustment for noninterviews in group quarters.)

The Bureau used complex techniques to adjust the weights for nonresponse. For a further explanation of the techniques used, see the <u>Nonresponse Adjustment Methods for Demographic Surveys at the U.S. Bureau of the Census</u>, November 1988, Working paper 8823, by R. Singh and R. Petroni. The success of these techniques in avoiding bias is unknown. An example of successfully avoiding bias can be found in "Current Nonresponse Research for the Survey of Income and Program Participation" (paper by Petroni, presented at the Second International Workshop on Household Survey Nonresponse, October 1991).

We performed an additional stage of adjustment to persons' weights to reduce the mean square errors of the survey estimates. We accomplished this by ratio adjusting the sample estimates to agree with monthly Current Population Survey (CPS) type estimates of the civilian (and some military) noninstitutional population of the United States at the national level by demographic characteristics including age, sex, and race as of the specified date. The Bureau brought CPS estimates by age, sex, and race into agreement with adjusted estimates from the 1990 decennial census. Adjustments to the 1990 decennial census estimates include an adjustment for undercount[1] and also reflect births, deaths, immigration, emigration, and changes in the Armed Forces since 1990. In addition, we controlled SIPP estimates to independent Hispanic controls and made an adjustment to assign equal weights to husbands and wives within the same household. We implemented all of the above adjustments for each reference month and the interview month.

The 1991 panel wave 6 is the first panel and wave to use the 1990 census based controls in the weighting. Weights for earlier waves were based on independent population estimates derived by updating the 1980 decennial census counts.

Tables 5 through 10 show the effect of the new population controls on:

- age,
- sex,
- race,
- Hispanic Origin,
- household type,
- mean monthly income,
- program participation,
- labor force participation, and
- health insurance coverage

by comparing the 1991 panel wave 6 estimates using 1990 census based population controls to estimates using the updated 1980 census based population controls. The 1990 decennial population counts differed somewhat from the independent estimate derived by updating the 1980 counts. The estimates show differences in the absolute numbers

---

[1] See "The 1990 Post-Enumeration Survey: Operations and Results" by Howard Hogan in the 1993 Proceedings of the Undercount in the 1990 Census Section, American Statistical Association.

such as 247,860,000 total nonfarm population based on the 1980 controls compared to 250,420,000 persons based on 1990 controls.

The use of the new controls may have a significant impact on the absolute numbers. However, this difference has little impact on the weighted survey estimates of summary measures (such as means and medians) and proportional measures (such as percent distributions). The distribution of households by type by race and Hispanic Origin are nearly identical, as are the distributions of persons by age by sex. The 1980 based and 1990 based estimates of mean household income were similar ($3,526 and $3,517, respectively). Also, the proportion of persons receiving benefits from means-tested programs (22.9 percent 1980 based compared to 23.3 percent 1990 based), the percent of persons with some labor force activity (66.2 percent 1980 based compared to 66.4 percent 1990 based), and the proportion of persons without any health insurance coverage (13.5 percent 1980 based compared to 13.7 percent 1990 based) did not show substantial differences between estimates based on different population controls.

Use of Weights. Each household and each person within each household on each wave tape has five weights. Four of these weights are reference month specific and therefore can be used only to form reference month estimates. Average reference month estimates to form estimates of monthly averages over some period of time. For example, using the proper weights, one can estimate the monthly average number of households in a specified income range over November and December 1990. To estimate monthly averages of a given measure (e.g., total, mean) over a number of consecutive months, sum the monthly estimates and divide by the number of months.

The remaining weight is interview month specific. Use this weight to form estimates that specifically refer to the interview month (e.g., total persons currently looking for work), as well as estimates referring to the time period including the interview month and all previous months (e.g., total persons who have ever served in the military).

To form an estimate for a particular month, use the reference month weight for the month of interest, summing over all persons or households with the characteristic of interest whose reference period includes the month of interest. Multiply the sum by a factor to account for the number of rotations contributing data for the month. This factor equals four divided by the number of rotations contributing data for the month. For example, December 1991 data is only available from rotations 2, 3, and 4 for Wave 1 of the 1991 panel (see table 3), so apply a factor of 4/3. To form an estimate for an interview month, use the procedure discussed above using the interview month weight provided on the file.

Apply factors greater than 1 when constructing estimates for months with four rotations worth of data from a wave file. However, when using core data from consecutive waves together, data from all four rotations may be available, in which case the factors are equal to 1.

These tapes contain no weight for characteristics that involve a persons's or household's status over two or more months (e.g., number of households with a 50 percent increase in income between November and December 1990).

Producing Estimates for Census Regions and States. The total estimate for a region is the sum of the state estimates in that region. Using this sample, estimates for individual states are subject to very high variance and are not recommended. The state codes on the file are primarily of use for linking respondent characteristics with appropriate contextual variables (e.g., state-specific welfare criteria) and for tabulating data by user-defined groupings of states.

Producing Estimates for the Metropolitan Population. For Washington, DC and 11 states, we identify metropolitan or non-metropolitan residence (variable H*-METRO). In 34 additional states, where the non-metropolitan population in the sample was small enough to present a disclosure risk, we recoded a fraction of the metropolitan sample to be indistinguishable from non-metropolitan cases (H*-METRO=2). In these states, therefore, the cases coded as metropolitan (H*-METRO=1) represent only a subsample of that population.

In producing state estimates for a metropolitan characteristic, multiply the individual, family, or household weights by the metropolitan inflation factor for that state, presented in table 11. (This inflation factor compensates for the subsampling of the metropolitan population and is 1.0 for the states with complete identification of the metropolitan population.)

The same procedure applies when creating estimates for particular identified MSA's or CMSA's--apply the factor appropriate to the state. For multi-state MSA's, use the factor appropriate to each state part. For example, to tabulate data for the Washington, DC-MD-VA MSA, apply the Virginia factor of 1.0521 to weights for residents of the Virginia part of the MSA; Maryland and DC residents require no modification to the weights (i.e., their factors equal 1.0).

In producing regional or national estimates of the metropolitan population, it is also necessary to compensate for the fact that we don't identify a metropolitan subsample within two states (Mississippi and West Virginia) and one state-group (North Dakota - South Dakota - Iowa). Thus, use factors in the right-hand column of table 11 for regional and national estimates. The results of regional and national tabulations of the metropolitan population will be biased slightly. However, less than one-half of one percent of the metropolitan population is not represented.

Producing Estimates for the Non-Metropolitan Population. State, regional, and national estimates of the non-metropolitan population cannot be computed directly, except for Washington, DC and the 11 states where the factor for state tabulations in table 11 is 1.0. In all other states, the cases identified as not in the metropolitan subsample

(METRO=2) are a mixture of non-metropolitan and metropolitan households. Only an indirect method of estimation is available: first compute an estimate for the total population, then subtract the estimates for the metropolitan population. The results of these tabulations will be slightly biased.

Combined Panel Estimates. Both the 1991 and 1990 panels provide data for October 1990-August 1992. Thus, obtain estimates for these time periods by combining the corresponding panels. However, since the Wave 1 questionnaire differs from the subsequent waves' questionnaire and since the procedures changed between the 1990 and 1991 panels, we recommend that estimates not be obtained by combining Wave 1 data of the 1991 panel with data from another panel. In this case, use the estimate obtained from either panel. Additionally, even for other waves, care should be taken when combining data from two panels since questionnaires for the two panels differ somewhat and since the length of time in sample for interviews from the two panels differ.

Obtain combined panel estimates either (1) by combining estimates derived separately for the two panels or (2) by first combining data from the two files and then producing an estimate.

1. Combining Separate Estimates

Combine corresponding estimates from two consecutive year panels to create joint estimates by using the formula

$$\hat{\mathcal{J}} = W\hat{\mathcal{J}}_1 + (1-W)\hat{\mathcal{J}}_2 \qquad \text{(A)}$$

$\hat{\mathcal{J}}$ = joint estimate (total, mean, proportion,etc);

$\hat{\mathcal{J}}_1$ = estimate from the earlier panel;

$\hat{\mathcal{J}}_2$ = estimate from the later panel;

W = weighting factor of the earlier panel.

To combine the 1990 and 1991 panels use a W value of 0.613 unless one of the panels contributes no information to the estimate. In that case, assign the panel contributing information a factor of 1. Assign the other a factor of zero.

2. Combining Data from Separate Files

Start by first creating a file containing the data from the two panel files. Apply the weighting factor, W, to the weight of each person from the earlier panel and apply (1-W) to the weight of each person from the later panel. Then produce estimates using the same methodology as used to obtain estimates from a single panel.

Illustration for computing combined panel estimate.

Suppose SIPP estimates for Wave 5, 1990 panel show there were 441,000 households with monthly December income above $6,000. Also, suppose SIPP estimates for Wave 2, 1991 panel show there were 435,000 households with monthly December income above $6,000. Using formula (A), the joint level estimate is

$$\hat{\jmath} = (0.613)(441{,}000) + (0.387)(435{,}000) = 439{,}000$$

ACCURACY OF ESTIMATES

We base SIPP estimates on a sample. The sample estimates may differ somewhat from the values obtained from administering a complete census using the same questionnaire, instructions, and enumerators. The difference occurs because with an estimate based on a sample survey two types of errors are possible: nonsampling and sampling. We can provide estimates of the magnitude of the SIPP sampling error, but this is not true of nonsampling error. The next few sections describe SIPP nonsampling error sources, followed by a discussion of sampling error, its estimation, and its use in data analysis.

Nonsampling Variability. We attribute nonsampling errors to many sources, they include:

- inability to obtain information about all cases in the sample,
- definitional difficulties,
- differences in the interpretation of questions,
- inability or unwillingness on the part of the respondents to provide correct information,
- inability to recall information,
- errors made in collection (e.g. recording or coding the data),
- errors made in processing the data,

- errors made in estimating values for missing data,
- biases resulting from the differing recall periods caused by the interviewing pattern used,
- undercoverage.

We used quality control and edit procedures to reduce errors made by respondents, coders and interviewers. More detailed discussions of the existence and control of nonsampling errors in the SIPP are in the SIPP Quality Profile.

Undercoverage in SIPP resulted from missed living quarters and missed persons within sample households. It is known that undercoverage varies with age, race, and sex. Generally, undercoverage is larger for males than for females and larger for Blacks than for Nonblacks. Ratio estimation to independent age-race-sex population controls partially corrects for the bias due to survey undercoverage. However, biases exist in the estimates when persons in missed households or missed persons in interviewed households have characteristics different from those of interviewed persons in the same age-race-sex group. Further, we didn't adjust the independent population controls for undercoverage in the Census.

A common measure of survey coverage is the coverage ratio, the estimated population before ratio adjustment divided by the independent population control. Table 12 shows CPS coverage ratios for age-sex-race groups for 1992. The CPS coverage ratios can exhibit some variability from month to month, but these are a typical set of coverage ratios. Other Census Bureau household surveys like the SIPP experience similar coverage.

Comparability with Other Estimates. Exercise caution when comparing data from this report with data from other SIPP publications or with data from other surveys. Comparability problems are from varying seasonal patterns for many characteristics, different nonsampling errors, and different concepts and procedures. Refer to the SIPP Quality Profile for known differences with data from other sources and further discussion.

Sampling Variability. Standard errors indicate the magnitude of the sampling error. They also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data. The standard errors mostly measure the variations that occurred by chance because we surveyed a sample rather than the entire population.

USES AND COMPUTATION OF STANDARD ERRORS

Confidence Intervals. The sample estimate and its standard error enable one to construct confidence intervals, ranges that would include the average result of all possible samples with a known probability. For example, if we selected all possible samples and surveyed each of these under essentially the same conditions and with the same sample design, and if we calculated an estimate and its standard error from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average estimate derived from all possible samples is or is not contained in any particular computed interval. However, for a particular sample, one can say with a specified confidence that the confidence interval includes the average estimate derived from all possible samples.

Hypothesis Testing. One may also use standard errors for hypothesis testing. Hypothesis testing is a procedure for distinguishing between population characteristics using sample estimates. The most common type of hypothesis tested is 1) the population characteristics are identical versus 2) they are different. One can perform tests at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

Unless noted otherwise, all statements of comparison in the report passed a hypothesis test at the 0.10 level of significance or better. This means that, for differences cited in the report, the estimated absolute difference between parameters is greater than 1.6 times the standard error of the difference.

To perform the most common test, compute the difference $X_A - X_B$, where X_A and X_B are sample estimates of the characteristics of interest. A later section explains how to derive an estimate of the standard error of the difference $X_A - X_B$. Let that standard error be s_{DIFF}. If $X_A - X_B$ is between -1.6 times s_{DIFF} and +1.6 times s_{DIFF}, no conclusion about the characteristics is justified at the 10 percent significance level. If, on the other hand, $X_A - X_B$ is smaller than -1.6 times s_{DIFF} or larger than +1.6 times s_{DIFF}, the observed difference is significant at the 10 percent level. In this event, it is commonly accepted practice to say that the characteristics are different. Of course, sometimes this conclusion will be wrong. When the characteristics are, in fact, the same, there is a 10 percent chance of concluding that they are different.

Note that as we perform more tests, more erroneous significant differences will occur. For example, at the 10 percent significance level, if we perform 100 independent hypothesis tests in which there are no real differences, it is likely that about 10 erroneous differences will occur. Therefore, interpret the significance of any single test cautiously.

Note Concerning Small Estimates and Small Differences. We show summary measures in the report only when the base is 200,000 or greater. Because of the large standard errors involved, there is little chance that estimates will reveal useful information when computed on a base smaller than 200,000. Also, nonsampling error in one or more of the small number of cases providing the estimate can cause large relative error in that particular estimate. We show estimated numbers, however, even though the relative standard errors of these numbers are larger than those for the corresponding percentages. We provide smaller estimates primarily to permit such combinations of the categories as serve each user's needs. Therefore, be careful in the interpretation of small differences since even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Standard Error Parameters and Tables and Their Use. Most SIPP estimates have greater standard errors than those obtained through a simple random sample because we sampled clusters of living quarters for the SIPP. To derive standard errors at a moderate cost and applicable to a wide variety of estimates, we made a number of approximations. We grouped estimates with similar standard error behavior and developed two parameters (denoted "a" and "b") to approximate the standard error behavior of each group of estimates. Because the actual standard error behavior was not identical for all estimates within a group, the standard errors we computed from these parameters provide an indication of the order of magnitude of the standard error for any specific estimate. These "a" and "b" parameters vary by characteristic and by demographic subgroup to which the estimate applies. Use base "a" and "b" parameters found in table 13 for 1991 panel estimates. Note that for estimates which include data for wave 5 and beyond multiply the "a" and "b" parameters by 1.09 to account for sample attrition.

The factors provided in table 14 when multiplied by the base parameters of table 13 for a given subgroup and type of estimate give the "a" and "b" parameters for that subgroup and estimate type for the specified reference period. For example, the base "a" and "b" parameters for total number of households are -0.0001005 and 9,286, respectively. For Wave 1 the factor for October 1990 is 4 since only 1 rotation month of data is available. So, the "a" and "b" parameters for total household income in October 1990 based on Wave 1 are -0.0004020 and 37,144, respectively. Also for Wave 1, the factor for the first quarter of 1991 is 1.2222 since 9 rotation months of data are available (rotations 1 and 4 provide 3 rotations months each, while rotations 2 and 3 provide 1 and 2 rotation months, respectively). So the "a" and "b" parameters for total number of households in the first quarter of 1991 are -0.00001228 and 11,349, respectively for Wave 1.

Use the "a" and "b" parameters to calculate the standard error for estimated numbers and percentages. Because the actual standard error behavior was not identical for all estimates within a group, the standard errors computed from these parameters provide an indication of the order of magnitude of the standard error for any specific estimate. The following sections give methods for using these parameter for computation of approximate standard errors.

For users who wish further simplification, we also provide general standard errors in tables 15 and 18. Note that you need to adjust these standard errors by a factor from table 13. The standard errors resulting from this simplified approach are less accurate. Methods for using these parameters and tables for computation of standard errors are given in the following sections.

For the 1990, 1991 combined panel parameters, multiply the parameters in table 13 by the appropriate factor from table 22. The factors provided in table 23 adjust parameters for the number of rotation months available for a given estimate. These factors, when multiplied by the combined panel parameters derived from table 13 for a given subgroup and type of estimate, give the "a" and "b" parameters for that subgroup and estimate type for the specified combined reference period.

Table 19 provides base "a" and "b" parameters for calculating 1991 topical module variances. Table 20 provides base "a" and "b" parameters for computing the 1990, 1991 combined panel topical module variances.

Described below are procedures for calculating standard errors for the types of estimates most commonly used. Note specifically that these procedures apply only to reference month estimates or averages of reference month estimates. Refer to the section "Use of Weights" for a more detailed discussion of the construction of estimates. We included stratum codes and half sample codes on the tapes so users can compute variances directly by methods such as balanced repeated replications (BRR). William G. Cochran provides a list of references discussing the application of this technique. (See Sampling Techniques, 3rd Ed., New York: John Wiley and Sons, 1977, p. 321.)

Standard errors of estimated numbers. Obtain the approximate standard error, s_x, of an estimated number of persons, households, families, unrelated individuals and so forth, in one of two ways. Both apply when data from all four rotations are used to make the estimate. However, only the second method should be used when less than four rotations of data are available for the estimate. Note that neither method should be applied to dollar values.

The standard error may be obtained by the use of the formula

$$s_x = fs \qquad (1)$$

where f is the appropriate "f" factor from table 13, and s is the standard error on the estimate obtained by interpolation from table 15 or 16. Alternatively, approximate s_x using the formula,

$$s_x = \sqrt{ax^2 + bx} \qquad (2)$$

from which we calculated the standard errors in tables 15 and 16. Here x is the size of the estimate and "a" and "b" are the parameters associated with the particular type of characteristic. Use of formula 2 will provide more accurate results than the use of formula 1.

Illustration.

Suppose SIPP estimates for Wave 1 of the 1991 panel show that there were 472,000 households with monthly household income above $6,000. The appropriate parameters and factor from table 13 and the appropriate general standard error from table 15 are

$a = -0.0001005 \qquad b = 9{,}286 \qquad f = 1.00 \qquad s = 66{,}000$

Using formula 1, the approximate standard error is

$s_x = 66{,}000$

Using formula 2, the approximate standard error is

$$\sqrt{(-0.0001005)(472{,}000)^2 + (9{,}286)(472{,}000)} = 66{,}000$$

Using the standard error based on formula 2, the approximate 90-percent confidence interval as shown by the data is from 366,000 to 578,000. Therefore, a conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90% of all samples.

Illustration for computing standard errors for combined panel estimates.

Suppose the combined SIPP estimate for total number of males in the 16+ Income and Labor Force for Wave 5, 1990 panel and Wave 2, 1991 panel was 92,398,000. The combined panel parameters for total males are obtained by multiplying the appropriate "a" and "b" values from table 13 by the appropriate factors from tables 22 and 23. The 1991 parameters and factors are a = -0.0001005, b = 9,286, g = 0.4163 and factor =

1.0000, respectively. Thus, the combined panel parameters are a = -0.0000418 and b = 3,866. Using formula 2, the approximate standard error is

$$s = \sqrt{(-0.0000418)(92{,}398{,}000)^2 + (3866)(92{,}398{,}000)} = 19{,}000$$

Standard Error of a Mean. Define a mean as the average quantity of some item (other than persons, families, or households) per person, family or household. For example, it could be the average monthly household income of females age 25 to 34. Use formulas below to approximate the standard error of a mean. Because of the approximations used in developing formula 3, an estimate of the standard error of the mean obtained from this formula will generally underestimate the true standard error. The formula used to estimate the standard error of a mean $\bar{x}$ is

$$s_{\bar{x}} = \sqrt{\left(\frac{b}{y}\right)s^2} \qquad (3)$$

where y is the size of the base, s^2 is the estimated population variance of the item and b is the parameter associated with the particular type of item.

Estimate the population variance s^2 by one of two methods. In both methods we assume x_i is the value of the item for unit i. (Unit may be person, family, or household). To use the first method, divide the range of values for the item into c intervals. The upper and lower boundaries of interval j are Z_{j-1} and Z_j, respectively. Place each unit into one of c groups such that $Z_{j-1} < x_i \leq Z_j$.

The estimated population variance, s^2, is given by the formula:

$$s^2 = \sum_{j=1}^{c} p_j m_j^2 - \bar{x}^2, \qquad (4)$$

where p_j is the estimated proportion of units in group j, and $m_j = (Z_{j-1} + Z_j)/2$. We assume the most representative value of the item in group j is m_j. If group c is open-ended, i.e., no upper interval boundary exists, then an approximate value for m_c is

$$m_c = \frac{3}{2} Z_{c-1}.$$

Compute the mean, $\bar{x}$, using the following formula:

$$\bar{x} = \sum_{j=1}^{c} p_j m_j .$$

In the second method, the estimated population variance is given by

$$s^2 = \frac{\sum_{i=1}^{n} w_i x_i^2}{\sum_{i=1}^{n} w_i} - \bar{x}^2 , \tag{5}$$

where there are n units with the item of interest and w_i is the final weight for unit i. Compute the mean, $\bar{x}$, using the formula

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} .$$

When forming combined estimates using formula (A) from the section on combined panel estimates, calculate s^2, given by formula (4), by forming a distribution for each panel. Divide the range of values for the item into intervals. Obtain combined estimates for each interval using formula (A). Apply formula (4) to the combined distribution. To calculate $\bar{x}$ and s^2 given by formula (5), replace x_i by Wx_i for x_i from the earlier panel and $(1-W)x_i$ for x_i from the later panel.

Illustration.

Suppose that based on Wave 1 data, the distribution of monthly cash income for persons age 25 to 34 during the month of January 1991 is given in table 21.

Using formula 4 and the mean monthly cash income of $2,530 the approximate population variance, s^2, is

$$s^2 = \left(\frac{1,371}{39,851}\right)(150)^2 + \left(\frac{1,651}{39,851}\right)(450)^2 + \ldots.. +$$

$$\left(\frac{1,493}{39,851}\right)(9,000)^2 - (2,530)^2 = 3,159,887.$$

Using formula 3, the appropriate base "b" parameter and factor from table 13, the estimated standard error of a mean $\bar{x}$ is

$$s_{\bar{x}} = \sqrt{\left(\frac{7,514}{39,851,000}\right)(3,159,887)} = \$24$$

Standard error of an aggregate. We define an aggregate as the total quantity of an item summed over all the units in a group. Approximate the standard error of an aggregate using formula 6.

Because of the approximations used in developing formula (6), it will generally underestimate the true standard error. Let y be the size of the base, s^2 be the estimated population variance of the item obtained using formula (4) or (5) and b be the parameter associated with the particular type of item. The standard error of an aggregate is:

$$s_x = \sqrt{(b)\ (y)\, s^2} \qquad (6)$$

Standard Errors of Estimated Percentages. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends on the size of the percentage and its base. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more, e.g., the percent of people employed is more reliable than the estimated number of people employed. When the numerator and denominator of the percentage have different parameters, use the parameter (and appropriate factor) of the numerator. If proportions are presented instead of percentages, note that the standard error of a proportion is equal to the standard error of the corresponding percentage divided by 100.

We commonly estimate two types of percentages. The first is the percentage of persons, families or households sharing a particular characteristic such as the percent of persons owning their own home. The second type is the percentage of money or some similar concept held by a particular group of persons or held in a particular form. Examples are the percent of total wealth held by persons with high income and the percent of total income received by persons on welfare.

For the percentage of persons, families, or households, calculate the approximate standard error, $s_{(x,p)}$, of an estimated percentage p using the formula

$$s_{(x,p)} = fs \tag{7}$$

when estimating p using data from all four rotations.

In this formula, f is the appropriate "f" factor from table 13 and s is the standard error of the estimate from table 17 or 18.

Alternatively, approximate it by the formula:

$$s_{(x,p)} = \sqrt{\frac{b}{x}\,(p)\,\,(100-p)} \tag{8}$$

from which we calculated the standard errors in tables 17 and 18. Here x is the size of the subclass of social units which is the base of the percentage, p is the percentage ($0<p<100$), and b is the parameter associated with the characteristic in the numerator. Using this formula gives more accurate results than using formula 7 above. Use this formula to estimate p for data with less than four rotations.

<u>Illustration.</u>

Suppose that, in the month of January 1991, 6.7 percent of the 16,812,000 persons in nonfarm households with a mean monthly household cash income of $4,000 to $4,999, were black. Using formula 8 and the "b" parameter of 10,110 from table 13 and a factor of 1 for the month of January 1991 from table 14, the approximate standard error is

$$\sqrt{\frac{10{,}110}{(16{,}812{,}000)}\,(6.7)\,\,(100-6.7)} = 0.61\ \textit{percent}$$

Consequently, the 90 percent confidence interval as shown by these data is from 5.7 to 7.7 percent.

Percentages of money require a more complicated formula. Estimate a percentage of money one of two ways. It may be the ratio of two aggregates:

$$p_I = 100 \ (X_A \ / \ X_N)$$

or it may be the ratio of two means with an adjustment for different bases:

$$p_I = 100 \ (\hat{p}_A \ \bar{X}_A \ / \ \bar{X}_N)$$

where x_A and x_N are aggregate money figures, $\bar{x}_A$ and $\bar{x}_N$ are mean money figures, and $\hat{p}_A$ is the estimated number in group A divided by the estimated number in group N. In either case, we estimate the standard error as

$$s_I = \sqrt{\left(\frac{\hat{p}_A\bar{x}_A}{\bar{x}_N}\right)^2 \left[\left(\frac{s_p}{\hat{p}_A}\right)^2 + \left(\frac{s_A}{\bar{x}_A}\right)^2 + \left(\frac{s_B}{\bar{x}_N}\right)^2\right]} \ , \qquad (9)$$

where s_p is the standard error of $\hat{p}_A$, s_A is the standard error of $\bar{x}_A$ and s_B is the standard error of $\bar{x}_N$. To calculate s_p, use formula 8. Calculate the standard errors of $\bar{x}_N$ and $\bar{x}_A$ using formula 3.

Note that there is frequently some correlation between $\hat{p}_A$, $\bar{x}_N$, and $\bar{x}_A$. Depending on the magnitude and sign of the correlations, the standard error will be over or underestimated.

<u>Illustration.</u>

Suppose that in January 1991, 9.8% of the households own rental property, the mean value of rental property is \$72,121, the mean value of assets is \$78,734, and the corresponding standard errors are 0.31%, \$5799, and \$2867. In total there are 86,790,000 households. Then, the percent of all household assets held in rental property is

$$= 100 \left((0.098) \frac{72121}{78734}\right) = 9.0\%$$

Using formula (9), the appropriate standard error is

$$s_I = \sqrt{\left(\frac{(0.098)(72121)}{78734}\right)^2 \left[\left(\frac{0.0031}{0.098}\right)^2 + \left(\frac{5799}{72121}\right)^2 + \left(\frac{2867}{78734}\right)^2\right]}$$

= 0.008

= 0.8%

Standard Error of a Difference. The standard error of a difference between two sample estimates, x and y, is approximately equal to

$$s_{(x-y)} = \sqrt{s_x^2 + s_y^2} \qquad (10)$$

where s_x and s_y are the standard errors of the estimates x and y.

The estimates can be numbers, percents, ratios, etc. The above formula assumes that the correlation coefficient between the characteristics estimated by x and y is zero. If the correlation is really positive (negative), then this assumption will tend to cause overestimates (underestimates) of the true standard error.

Illustration.

Suppose that SIPP estimates show the number of persons age 35-44 years with monthly cash income of $4,000 to $4,999 was 3,186,000 in the month of January 1991 and the number of persons age 25-34 years with monthly cash income of $4,000 to $4,999 in the same time period was 2,619,000. Then, using parameters from table 13 and formula 2, the standard errors of these numbers are approximately 153,000 and 139,000, respectively. The difference in sample estimates is 567,000 and, using formula 10, the approximate standard error of the difference is

$$\sqrt{(153{,}000)^2 + (139{,}000)^2} = 207{,}000$$

Suppose that it is desired to test at the 10 percent significance level whether the number of persons with monthly cash income of $4,000 to $4,999 was different for persons age 35-44 years than for persons age 25-34 years. To perform the test, compare the difference of 567,000 to the product 1.6 x 207,000 = 331,200. Since the difference is greater than 1.6 times the standard error of the difference, the data show that the two age groups are significantly different at the 10 percent significance level.

Standard Error of a Median. The median quantity of some item such as income for a given group of persons, families, or households is that quantity such that at least half the group have as much or more and at least half the group have as much or less. The sampling variability of an estimated median depends upon the form of the distribution of the item as well as the size of the group. Use the procedure described below to calculate standard errors on medians.

An approximate method for measuring the reliability of an estimated median is to determine a confidence interval about it. (See the section on sampling variability for a general discussion of confidence intervals.) Use the following procedure to estimate the 68-percent confidence limits and hence the standard error of a median based on sample data.

1. Determine, using either formula 7 or formula 8, the standard error of an estimate of 50 percent of the group;

2. Add to and subtract from 50 percent the standard error determined in step 1;

3. Using the distribution of the item within the group, calculate the quantity of the item such that the percent of the group with more of the item is equal to the smaller percentage found in step 2. This quantity will be the upper limit for the 68-percent confidence interval. In a similar fashion, calculate the quantity of the item such that the percent of the group with more of the item is equal to the larger percentage found in step 2. This quantity will be the lower limit for the 68-percent confidence interval;

4. Divide the difference between the two quantities determined in step 3 by two to obtain the standard error of the median.

To perform step 3, you must interpolate. You may use different methods of interpolation. The most common are simple linear interpolation and Pareto interpolation. The appropriateness of the method depends on the form of the distribution around the median. If density is declining in the area, then we recommend Pareto interpolation. If density is fairly constant in the area, then we recommend linear interpolation. Never use Pareto interpolation if the interval contains zero or negative measures of the item of interest. Use interpolation as follows. The quantity of the item such that "p" percent have more of the item is

$$X_{pN} = \exp\left[\left(Ln\left(\frac{pN}{N_1}\right) / Ln\left(\frac{N_2}{N_1}\right)\right) Ln\left(\frac{A_2}{A_1}\right)\right] A_1 \qquad (11)$$

if Pareto Interpolation is indicated and

$$X_{pN} = \left[\frac{PN-N_1}{N_2-N_1} \quad (A_2-A_1) + A_1\right] \tag{12}$$

if linear interpolation is indicated, where

N	is the size of the group,
A_1 and A_2	are the lower and upper bounds, respectively, of the interval in which X_{pN} falls,
N_1 and N_2	are the estimated number of group members owning more than A_1 and A_2, respectively,
exp	refers to the exponential function and
Ln	refers to the natural logarithm function.

<u>Illustration.</u>

To illustrate the calculations for the sampling error on a median, we return to table 21. The median monthly income for this group is $2,158. The size of the group is 39,851,000.

1. Using formula 8, the standard error of 50 percent on a base of 39,851,000 is about 0.7 percentage points.

2. Following step 2, the two percentages of interest are 49.3 and 50.7.

3. By examining table 21, we see that the percentage 49.3 falls in the income interval from 2000 to 2499. (Since 55.5% receive more than $2,000 per month, the dollar value corresponding to 49.3 must be between $2,000 and $2,500). Thus, A_1 = $2,000, A_2 = $2,500, N_1 = 22,106,000, and N_2 = 16,307,000.

In this case, we decided to use Pareto interpolation. Therefore, the upper bound of a 68% confidence interval for the median is

$$\$2{,}000 \exp\left[\left(Ln\left(\frac{(.493)\,(39{,}851{,}000)}{22{,}106{,}000}\right) / Ln\left(\frac{16{,}307{,}000}{22{,}106{,}000}\right)\right) Ln\left(\frac{2{,}500}{2{,}000}\right)\right] = \$2181$$

Also by examining table 21, we see that 50.7 falls in the same income interval. Thus, A_1, A_2, N_1 and N_2 are the same. We also use Pareto interpolation for this case. So the lower bound of a 68% confidence interval for the median is

$$\$2,000 \exp\left[\left(Ln\left(\frac{(.507)(39,851,000)}{22,106,000}\right) \Big/ Ln\left(\frac{16,307,000}{22,106,000}\right)\right) Ln\left(\frac{2,500}{2,000}\right)\right] = \$2136$$

Thus, the 68-percent confidence interval on the estimated median is from \$2136 to \$2181. An approximate standard error is

$$\frac{\$2181 - \$2136}{2} = \$23$$

Standard Errors of Ratios of Means and Medians. Approximate the standard error for a ratio of means or medians by:

$$s_{\frac{x}{y}} = \sqrt{\left(\frac{x}{y}\right)^2 \left[\left(\frac{s_y}{y}\right)^2 + \left(\frac{s_x}{x}\right)^2\right]} \qquad (13)$$

where x and y are the means or medians, and s_x and s_y are their associated standard errors. Formula 13 assumes that the means are not correlated. If the correlation between the population means estimated by x and y are actually positive (negative), then this procedure will tend to produce overestimates (underestimates) of the true standard error for the ratio of means.

Table 1. 1991 Panel Topical Modules

Wave	Topical Module
1	None
2	Recipiency History Employment History Work Disability History Education and Training History Marital History Migration History Fertility History Household Relationships
3	Child Care Arrangements Child Support Agreements Support of Non-household Members Functional Limitations and Disability Utilization of Health Care Services Work Schedule
4	Selected Financial Assets Medical Expenses and Work Disability Real Estate, Shelter Costs, Dependent Care, and Vehicles
5	Taxes Annual Income and Retirement Accounts School Enrollment and Financing
6	Extended Measures of Wellbeing (Consumer Durables, Living Conditions, Basic Needs, Expenditures, Minimum Income)
7	Assets and Liabilities Retirement Expectations and Pension Plan Coverage Real Estate Property and Vehicles
8	Taxes Annual Income and Retirement Accounts School Enrollment and Financing

Table 2. 1990 Panel Topical Modules

Wave	Topical Module
1	None
2	Recipiency History Employment History Work Disability History Education and Training History Marital History Migration History Fertility History Household Relationships
3	Work Schedule Child Care Child Support Agreements Support of Non-household Members Functional Limitations and Disability Utilization of Health Care Services
4	Assets and Liabilities Retirement Expectations and Pension Plan Coverage Real Estate Property and Vehicles
5	Taxes Annual Income and Retirement Accounts School Enrollment and Financing
6	Child Support Agreements Support for Non-household Members Functional Limitations and Disability Utilization of Health Care Services Not in Labor Force Spells
7	Selected Financial Assets Medical Expenses and Work Disability Real Estate, Shelter Costs, Dependent Care and Vehicles
8	Taxes Annual Income and Retirement Accounts School Enrollment and Financing

Table 3. Reference Months for Each Interview Month - 1991 Panel

Reference Period

		4th Quarter (1990)			1st Quarter (1991)			2nd Quarter (1991)			3rd Quarter (1991)			4th Quarter (1991)			...	2nd Quarter (1993)			3rd Quarter (1993)		
Month of Interview	Wave/ Rotation	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Apr	May	Jun	Jul	Aug	Sep
Feb 91	1/2	X	X	X	X																		
Mar	1/3		X	X	X	X																	
Apr	1/4			X	X	X	X																
May	1/1				X	X	X	X															
Jun	2/2					X	X	X	X														
Jul	2/3						X	X	X	X													
Aug	2/4							X	X	X	X												
Sept	2/1								X	X	X	X											
Oct	3/2									X	X	X	X										
Nov	3/3										X	X	X	X									
Dec	3/4											X	X	X	X								
.														.	.	.							
.																.	...						
.																	...	.					
Sept 93	8/1																		X	X	X	X	

Table 4. Reference Months for Each Interview Month - 1990 Panel

		Reference Period																					
		4th Quarter (1989)			1st Quarter (1990)			2nd Quarter (1990)			3rd Quarter (1990)			4th Quarter (1990)			...	2nd Quarter (1992)			3rd Quarter (1992)		
Month of Interview	Wave/ Rotation	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Apr	May	Jun	Jul	Aug	Sep
Feb 90	1/2	X	X	X	X																		
Mar	1/3		X	X	X	X																	
Apr	1/4			X	X	X	X																
May	1/1				X	X	X	X															
Jun	2/2					X	X	X	X														
Jul	2/3						X	X	X	X													
Aug	2/4							X	X	X	X												
Sept	2/1								X	X	X	X											
Oct	3/2									X	X	X	X										
Nov	3/3										X	X	X	X									
Dec	3/4											X	X	X	X								
.														.	.	.							
.																.	...						
.																	...	.					
Sept 92	8/1																		X	X	X	X	

Table 5. Non-Farm Population by Age and Sex: 1991 Panel Wave 6

AGES	Based on 1980 census population controls						Based on 1990 census population controls					
	Total		Males		Females		Total		Males		Females	
	Number (thous.)	Distribution	Number (thous.)	Distribution	Number (thous.)	Distribution	Number (thous.)	Distribution	Number (thous.)	Distribution	Number (thous.)	Distribution
All Ages	247864	100	120730	100	127134	100	250419	100	122128	100	128292	100
Under 4 years old	19270	7.8	9813	8.1	9457	7.4	19749	7.9	10054	8.2	9695	7.6
5 to 9 years old	18568	7.5	9585	7.9	8983	7.1	18898	7.5	9757	8.0	9142	7.1
10 to 15 years old	21207	8.6	10763	8.9	10444	8.2	21720	8.7	11014	9.0	10706	8.3
16 to 24 years old	30450	12.3	15231	12.6	15219	12.0	32156	12.8	16201	13.3	15954	12.4
25 to 34 years old	41571	16.8	20619	17.1	20952	16.5	42013	16.8	20818	17.0	21195	16.5
35 to 44 years old	39163	15.8	19331	16.0	19832	15.6	39536	15.8	19540	16.0	19996	15.6
45 to 54 years old	27075	10.9	13146	10.9	13929	11.0	26763	10.7	13073	10.7	13690	10.7
55 to 64 years old	20128	8.1	9518	7.9	10610	8.3	19708	7.9	9334	7.6	10374	8.1
65 to 69 years old	9972	4.0	4564	3.8	5408	4.3	9673	3.9	4371	3.6	5302	4.1
70 to 74 years old	8013	3.2	3454	2.9	4559	3.6	7878	3.1	3347	2.7	4532	3.5
75 years old and over	12446	5.0	4706	3.9	7740	6.1	12325	4.9	4619	3.8	7706	6.0

Table 6. Household Composition by Race and Hispanic Origin: 1991 Panel Wave 6

CHARACTERISTICS	Based on 1980 census population controls								Based on 1990 census population controls							
	All Races		White		Black		Hispanic Origin		All Races		White		Black		Hispanic Origin	
	Number (tens of thous.)	Dist.	Number (tens of thous.)	Dist.	Number (tens of thous.)	Dist.	Number (tens of thous.)	Dist.	Number (tens of thous.)	Dist.	Number (tens of thous.)	Dist.	Number (tens of thous.)	Dist.	Number (tens of thous.)	Dist.
All households	9610	100	8206	100	1099	100	677	100	9601	100	8190	100	1106	100	730	100
Family households	6758	70.3	5759	70.2	759	69.1	531	78.4	6760	70.4	5753	70.2	768	69.4	570	78.1
With own children under 18	3284	34.2	2711	33	429	39	335	49.5	3320	34.6	2731	33.3	443	40.1	363	49.7
Married-couple family	5290	55	4736	57.7	368	33.5	360	53.2	5280	55	4724	57.7	371	33.5	384	52.6
With own children under 18	2459	25.6	2159	26.3	187	17	232	34.3	2480	25.8	2173	26.5	193	17.5	249	34.1
Female householder	1175	12.2	779	9.5	357	32.5	137	20.2	1185	12.3	783	9.6	362	32.7	149	20.4
With own children under 18	711	7.4	454	5.5	231	21	90	13.3	723	7.5	460	5.6	238	21.5	98	13.4
Male householder	293	3	244	3	34	3.1	34	5	295	3.1	246	3	35	3.2	37	5.1
With own children under 18	114	1.2	98	1.2	11	1	13	1.9	117	1.2	98	1.2	12	1.1	16	2.2
Non family households	2851	29.7	2447	29.8	340	30.9	146	21.6	2841	29.6	2438	29.8	338	30.6	160	21.9
Living alone	2473	25.7	2119	25.8	302	27.5	121	17.9	2456	25.6	2106	25.7	299	27	132	18.1
Male householder	1252	13	1062	12.9	156	14.2	73	10.8	1250	13	1060	12.9	157	14.2	82	11.2
Living alone	1019	10.6	864	10.5	128	11.6	55	8.1	1013	10.6	859	10.5	127	11.5	62	8.5
Female householder	1600	16.6	1385	16.9	184	16.7	73	10.8	1590	16.6	1378	16.8	181	16.4	78	10.7
Living alone	1454	15.1	1255	15.3	174	15.8	66	9.7	1443	15	1247	15.2	172	15.6	70	9.6

Table 7. Selected Characteristics of Persons, by Mean Monthly Household Cash Income: Monthly Average for 1991 Panel Wave 6.

CHARACTERISTICS	Based on 1980 census population controls			Based on 1990 census population controls		
	Total (thous.)	Mean monthly cash income (dols.)		Total (thous.)	Mean monthly cash income (dols.)	
		Value	Standard error		Value	Standard error
Total	247,860	3,526	116	250,420	3,517	115
RACE AND HISPANIC ORIGIN						
White	205,980	3,670	130	207,960	3,659	129
Black	31,710	2,361	124	32,210	2,366	124
Hispanic origin	22,180	2,573	130	25,000	2,568	122
AGE						
Under 16 years old	59,050	3,332	221	60,370	3,308	218
16 to 24 years old	30,450	3,772	365	32,160	3,757	353
25 to 34 years old	41,570	3,441	230	42,010	3,432	228
35 to 44 years old	39,160	3,998	297	39,540	3,987	296
45 to 54 years old	27,080	4,443	420	26,760	4,438	422
55 to 64 years old	20,130	3,609	506	19,710	3,612	512
65 years old and over	30,430	2,293	221	29,880	2,291	223
EDUCATION						
25 years and over	158,370	3,551	145	157,900	3,548	145
Elem.: Less than 8 years	9,740	2,204	521	9,750	2,203	518
8 years	6,280	1,923	379	6,240	1,924	381
High School: 1 to 3 years	18,390	2,257	294	18,310	2,256	295
4 years	58,630	3,163	182	58,400	3,161	182
College 1 to 3 years	30,550	3,732	310	30,550	3,729	310
4 years	18,980	5,081	535	18,930	5,075	535
5 years or more	15,790	5,788	641	15,700	5,783	642
REGION						
Northeast	51,660	3,842	283	52,030	3,836	282
Midwest	62,650	3,547	222	62,790	3,547	222
South	80,100	3,106	181	81,050	3,098	179
West	53,460	3,825	264	54,560	3,801	260

Table 8. Selected Characteristics of Persons, By Program Participation Status: Monthly Average for 1991 Panel Wave 6.

CHARACTERISTICS	Based on 1980 census population controls							Based on 1990 census population controls						
	Total (thous.)	Residing in household receiving one or more means-tested program						Total (thous.)	Residing in a household receiving one or more means-tested program					
		Total		Cash benefit		Noncash benefit			Total		Cash benefit		Noncash benefit	
		Number	Percent of total	Number	Percent of total	Number	Percent of total		Number	Percent of total	Number	Percent of total	Number	Percent of total
Total	247,860	56,820	22.9	25,610	10.3	5,602	22.6	250,420	58,350	23.3	26,220	10.5	57,550	23.0
RACE AND HISPANIC ORIGIN														
White	205,980	37,770	18.3	14,300	6.9	37,230	18.1	207,960	38,940	18.7	14,720	7.1	38,400	18.5
Black	31,710	15,840	50.0	9,630	30.4	15,600	49.2	32,210	16,170	50.2	9,810	30.5	15,930	49.5
Hispanic origin	22,180	10,490	47.3	4,460	20.1	10,430	47.0	25,000	11,900	47.6	5,050	20.2	11,840	47.4
AGE														
Under 16 years old	59,050	21,550	36.5	9,140	15.5	21,490	36.4	60,370	22,370	37.1	9,500	15.7	22,310	37.0
16 to 24 years old	30,450	7,660	25.2	3,540	11.6	7,610	25.0	32,160	8,200	25.5	3,780	11.8	8,140	25.3
25 to 34 years old	41,570	9,350	22.5	3,570	8.6	9,280	22.3	42,010	9,520	22.7	3,620	8.6	9,460	22.5
35 to 44 years old	39,160	6,890	17.6	2,810	7.2	6,800	17.4	39,540	7,040	17.8	2,870	7.3	6,950	17.6
45 to 54 years old	27,080	3,340	12.3	1,920	7.1	3,250	12.0	26,760	3,320	12.4	1,900	7.1	3,240	12.1
55 to 64 years old	20,130	2,660	13.2	1,580	7.9	2,530	12.6	19,710	2,610	13.3	1,550	7.9	2,480	12.6
65 years old and over	30,430	5,370	17.6	3,050	10.0	5,060	16.6	29,880	5,270	17.7	2,990	10.0	4,980	16.7
EDUCATION														
25 years and over	158,370	27,610	17.4	12,920	8.2	26,920	17.0	157,900	27,780	17.6	12,940	8.2	27,100	17.2
Elem.: Less than 8 years	9,740	4,170	42.8	2,450	25.2	4,060	41.7	9,750	4,200	43.1	2,450	25.1	4,100	42.1
8 years	6,280	1,890	30.0	1,020	16.3	1,800	28.6	6,240	1,890	30.3	1,020	16.3	1,800	28.9
High School: 1 to 3 years	18,390	5,590	30.4	3,030	16.5	5,450	29.6	18,310	5,620	30.7	3,040	16.6	5,480	29.9
4 years	58,630	10,150	17.3	4,200	7.2	9,920	16.9	58,400	10,200	17.5	4,200	7.2	9,980	17.1
College 1 to 3 years	30,550	3,750	12.3	1,490	4.9	3,690	12.1	30,550	3,790	12.4	1,500	4.9	3,730	12.2
4 years	18,980	1,240	6.5	490	2.6	1,200	6.3	18,930	1,250	6.6	490	2.6	1,210	6.4
5 years or more	15,790	820	5.2	240	1.5	800	5.1	15,700	830	5.3	240	1.5	810	5.1
REGION														
Northeast	51,660	10,940	21.2	5,340	10.3	10,840	21.0	52,030	11,180	21.5	5,510	10.6	11,080	21.3
North Central	62,650	11,390	18.2	5,140	8.2	11,100	17.7	62,790	11,510	18.3	5,210	8.3	11,220	17.9
South	80,100	21,530	26.9	9,200	11.5	21,200	26.5	81,050	22,130	27.3	9,400	11.6	21,810	26.9
West	53,460	12,960	24.2	5,930	11.1	12,890	24.1	54,560	13,530	24.8	6,110	11.2	13,450	24.7

Table 9. Selected Characteristics of Persons, by Labor Force Status: Monthly Average for 1991 Panel Wave 6

LABOR FORCE ACTIVITY, AGE, AND SEX	Based on 1980 census population controls		Based on 1990 census population controls	
	Number (thous.)	Distribution	Number (thous.)	Distribution
BOTH SEXES				
Total, 16 years and over	188,819	100.0	190,053	100.0
With some labor force activity	124,945	66.2	126,127	66.4
With job entire month	114,431	60.6	115,349	60.7
Worked each week	111,399	59.0	112,298	59.1
Full-time worker	90,796	48.1	91,449	48.1
Part-time worker	20,603	10.9	20,850	11.0
Absent one or more weeks	3,032	1.6	3,051	1.6
With job part of month	2,717	1.4	2,783	1.5
Spent time looking or on layoff	1,364	0.7	1,399	0.7
No job during month	7,797	4.1	7,994	4.2
Looking for work or on layoff entire month	7,142	3.8	7,320	3.9
Looking for work or on layoff part of month	655	0.3	674	0.4
With no labor force activity	63,874	33.8	63,926	33.6
MALE				
Total, 16 years and over	90,569	100.0	91,304	100.0
With some labor force activity	67,716	74.8	68,516	75.0
With job entire month	61,818	68.3	62,456	68.4
Worked each week	60,535	66.8	61,158	67.0
Full-time worker	53,714	59.3	54,195	59.4
Part-time worker	6,821	7.5	6,962	7.6
Absent one or more weeks	1,284	1.4	1,298	1.4
With job part of month	1,373	1.5	1,415	1.5
Spent time looking or on layoff	788	0.9	811	0.9
No job during month	4,524	5.0	4,645	5.1

Table 9. cont'd Selected Characteristics of Persons, by Labor Force Status: Monthly Average for 1991 Panel Wave 6

LABOR FORCE ACTIVITY, AGE, AND SEX	Based on 1980 census population controls		Based on 1990 census population controls	
	Number (thous.)	Distribution	Number (thous.)	Distribution
Looking for work or on layoff entire month	4,286	4.7	4,399	4.8
Looking for work or on layoff part of month	238	0.3	246	0.3
With no labor force activity	22,853	25.2	22,788	25.0
FEMALE				
Total, 16 years and over	98,250	100.0	98,749	100.0
With some labor force activity	57,229	58.2	57,611	58.3
With job entire month	52,613	53.6	52,894	53.6
Worked each week	50,865	51.8	51,141	51.8
Full-time worker	37,082	37.7	37,253	37.7
Part-time worker	13,782	14.0	13,887	14.1
Absent one or more weeks	1,748	1.8	1,753	1.8
With job part of month	1,343	1.4	1,368	1.4
Spent time looking or on layoff	576	0.6	588	0.6
No job during month	3,273	3.3	3,349	3.4
Looking for work or on layoff entire month	2,856	2.9	2,920	3.0
Looking for work or on layoff part of month	416	0.4	429	0.4
With no labor force activity	41,021	41.8	41,138	41.7

Table 10. Selected Characteristics of Persons, by Health Insurance Coverage: Monthly Average for 1991 Panel Wave 6.

CHARACTERISTICS	Based on 1980 census population control							Based on 1990 census population control						
	Total (thous.)	Covered by private or government health insurance				Not covered by private or government health insurance		Total (thous.)	Covered by private or government health insurance				Not covered by private or government health insurance	
		Number	Percent of total	Covered by private health insurance					Number	Percent of total	Covered by private health insurance			
				Number	Percent of total	Number	Percent of total				Number	Percent of total	Number	Percent of total
Total	253,050	218,940	86.5	188,780	74.6	34,110	13.5	255,610	220,500	86.3	189,830	74.3	35,110	13.7
RACE AND HISPANIC ORIGIN														
White	210,980	184,610	87.5	164,730	78.1	26,370	12.5	212,960	185,740	87.2	165,440	77.7	27,220	12.8
Black	31,800	25,880	81.4	17,090	53.7	5,920	18.6	32,300	26,260	81.3	17,390	53.8	6,040	18.7
Hispanic origin	22,380	16,100	71.9	11,470	51.3	6,280	28.1	25,220	18,070	71.6	12,850	51.0	7,150	28.4
AGE														
Under 16 years old	60,170	52,250	86.8	41,390	68.8	7,920	13.2	61,490	53,240	86.6	41,970	68.3	8,250	13.4
16 to 24 years old	31,120	24,540	78.9	21,660	69.6	6,580	21.1	32,860	25,800	78.5	22,720	69.1	7,060	21.5
25 to 34 years old	42,160	34,060	80.8	30,510	72.4	8,100	19.2	42,600	34,370	80.7	30,770	72.2	8,230	19.3
35 to 44 years old	39,950	34,260	85.8	31,770	79.5	5,690	14.2	40,320	34,530	85.6	31,990	79.3	5,790	14.4
45 to 54 years old	27,770	24,360	87.7	22,660	81.6	3,420	12.3	27,450	24,050	87.6	22,370	81.5	3,400	12.4
55 to 64 years old	20,820	18,610	89.4	16,840	80.9	2,210	10.6	20,390	18,220	89.4	16,470	80.8	2,180	10.7
65 years old and over	31,060	30,860	99.4	23,940	77.1	200	0.6	30,490	30,300	99.4	23,530	77.2	200	0.7
REGION														
Northeast	52,080	46,700	89.7	40,310	77.4	5,380	10.3	52,440	46,940	89.5	40,440	77.1	5,500	10.5
North Central	65,570	59,080	90.1	53,530	81.6	6,480	9.9	65,700	59,140	90.0	53,540	81.5	6,560	10.0
South	81,460	67,600	83.0	56,370	69.2	13,850	17.0	82,410	68,160	82.7	56,780	68.9	14,250	17.3
West	53,950	45,550	84.4	38,560	71.5	8,400	15.6	55,060	46,270	84.0	39,070	71.0	8,790	16.0

Table 11. Metropolitan Subsample Factors to be Applied to Compute National and Subnational Estimates

		Factors for use in State or CMSA (MSA) Tabulations	Factors for use in Regional or National Tabulations
Northeast:	Connecticut	1.0387	1.0387
	Maine	1.2219	1.2219
	Massachusetts	1.0000	1.0000
	New Hampshire	1.2234	1.2234
	New Jersey	1.0000	1.0000
	New York	1.0000	1.0000
	Pennsylvania	1.0096	1.0096
	Rhode Island	1.2506	1.2506
	Vermont	1.2219	1.2219
Midwest:	Illinois	1.0000	1.0110
	Indiana	1.0336	1.0450
	Iowa	---	---
	Kansas	1.2912	1.3055
	Michigan	1.0328	1.0442
	Minnesota	1.0366	1.0480
	Missouri	1.0756	1.0874
	Nebraska	1.6289	1.6468
	North Dakota	---	---
	Ohio	1.0233	1.0346
	South Dakota	---	---
	Wisconsin	1.0188	1.0300
South:	Alabama	1.1574	1.1595
	Arkansas	1.6150	1.6179
	Delaware	1.5593	1.5621
	D.C.	1.0000	1.0018
	Florida	1.0140	1.0158
	Georgia	1.0142	1.0160
	Kentucky	1.2120	1.2142
	Louisiana	1.0734	1.0753
	Maryland	1.0000	1.0018
	Mississippi	---	---
	North Carolina	1.0000	1.0018
	Oklahoma	1.0793	1.0812
	South Carolina	1.0185	1.0203
	Tennessee	1.0517	1.0536
	Texas	1.0113	1.0131
	Virginia	1.0521	1.0540
	West Virginia	---	---

- indicates no metropolitan subsample is identified for the state

Table 11 cont'd. Metropolitan Subsample Factors to be Applied to Compute National and Subnational Estimates

		Factors for use in State or CMSA (MSA) Tabulations	Factors for use in Regional or National Tabulations
West:	Alaska	1.4339	1.4339
	Arizona	1.0117	1.0117
	California	1.0000	1.0000
	Colorado	1.1306	1.1306
	Hawaii	1.0000	1.0000
	Idaho	1.4339	1.4339
	Montana	1.4339	1.4339
	Nevada	1.0000	1.0000
	New Mexico	1.0000	1.0000
	Oregon	1.1317	1.1317
	Utah	1.0000	1.0000
	Washington	1.0456	1.0456
	Wyoming	1.4339	1.4339

- indicates no metropolitan subsample is identified for the state

Table 12. 1991 CPS Coverage Ratios

Age	non-Black		Black		All Persons		
	Male	Female	Male	Female	Male	Female	Total
0-14	0.963	0.965	0.927	0.926	0.957	0.959	0.958
15	0.962	0.949	0.899	0.919	0.952	0.944	0.948
16	0.969	0.936	0.923	0.907	0.962	0.932	0.947
17	0.981	0.975	0.945	0.862	0.975	0.957	0.966
18	0.939	0.926	0.883	0.846	0.930	0.913	0.922
19	0.860	0.872	0.754	0.801	0.844	0.861	0.853
20-24	0.913	0.927	0.734	0.832	0.889	0.913	0.901
25-26	0.927	0.940	0.688	0.877	0.897	0.931	0.914
27-29	0.910	0.954	0.707	0.864	0.885	0.941	0.914
30-34	0.893	0.948	0.691	0.883	0.870	0.939	0.905
35-39	0.910	0.949	0.763	0.899	0.895	0.942	0.919
40-44	0.929	0.951	0.824	0.906	0.919	0.946	0.933
45-49	0.956	0.966	0.903	0.956	0.951	0.965	0.958
50-54	0.940	0.961	0.807	0.877	0.927	0.951	0.940
55-59	0.944	0.941	0.826	0.825	0.932	0.928	0.930
60-62	0.965	0.956	0.792	0.850	0.948	0.944	0.946
63-64	0.905	0.907	0.669	0.872	0.884	0.903	0.894
65-67	0.935	0.979	0.783	0.875	0.921	0.969	0.947
68-69	0.925	0.942	0.789	0.831	0.913	0.931	0.923
70-74	0.926	0.993	0.856	1.014	0.920	0.995	0.962
75-99	0.977	0.989	0.764	0.912	0.961	0.983	0.975
15+	0.928	0.953	0.782	0.883	0.912	0.944	0.929
0+	0.936	0.955	0.827	0.895	0.923	0.947	0.935

Table 13: SIPP Indirect Generalized Variance Parameters for the 1991 Panel

Characteristics[1]	Parameters		
	a	b	f
PERSONS			
Total or White			
16+ Program Participation and Benefits, Poverty (3)			
Both Sexes	-0.0001342	22,040	0.90
Male	-0.0002789	22,040	
Female	-0.0002587	22,040	
16+ Income and Labor Force (5)			
Both Sexes	-0.0000407	7,514	0.52
Male	-0.0000850	7,514	
Female	-0.0000778	7,514	
16+ Pension Plan[2] (4)			
Both Sexes	-0.0000744	13,761	0.71
Male	-0.0001556	13,761	
Female	-0.0001425	13,761	
All Others[2] (6)			
Both Sexes	-0.0001134	27,327	1.00
Male	-0.0002334	27,327	
Female	-0.0002203	27,327	
Black			
Poverty (1)			
Both Sexes	-0.0006397	18,800	0.83
Male	-0.0013668	18,800	
Female	-0.0012028	18,800	
All Others (2)			
Both Sexes	-0.0003441	10,110	0.61
Male	-0.0007350	10,110	
Female	-0.0006468	10,110	
HOUSEHOLDS			
Total or White	-0.0001005	9,286	1.00
Black	-0.0006115	6,416	0.83

---

[1] To account for sample attrition, multiply the a and b parameters by 1.09 for estimates which include data from Wave 5 and beyond.

For cross-tabulations, use the parameters of the characteristic with the smaller number within the parentheses.

[2] Use the "16+ Pension Plan" parameters for pension plan tabulations of persons 16+ in the labor force. Use the "All Others" parameters for retirement tabulations, 0+ program participation, 0+ benefits, 0+ income, and 0+ labor force tabulations, in addition to any other types of tabulations not specifically covered by another characteristic in this table.

Table 14. Factors to be Applied to Table 13 Base Parameters to Obtain Parameters for Various Reference Periods

# of available rotation months[1]	factor
Monthly estimate	
1	4.0000
2	2.0000
3	1.3333
4	1.0000
Quarterly estimate	
6	1.8519
8	1.4074
9	1.2222
10	1.0494
11	1.0370
12	1.0000

[1] The number of available rotation months for a given estimate is the sum of the number of rotations available for each month of the estimate.

Table 15. Standard Errors of Estimated Numbers of Households, Families or Unrelated Persons (Numbers in Thousands)

Size of Estimate	Standard Error[1]	Size of Estimate	Standard Error[1]
200	43	15,000	342
300	53	25,000	412
500	68	30,000	434
750	83	40,000	459
1,000	96	50,000	462
2,000	135	60,000	442
3,000	164	70,000	397
5,000	210	80,000	316
7,500	253	90,000	147
10,000	288	92,000	61

1 To account for sample attrition, multiply the standard error of the estimate by 1.04 for estimates which include data from Wave 5 and beyond.

Table 16. Standard Errors of Estimated Numbers of Persons (Numbers in Thousands)

Size of Estimate	Standard Error	Size of Estimate	Standard Error
200	74	50,000	1041
300	90	80,000	1208
600	128	100,000	1264
1,000	165	130,000	1279
2,000	233	135,000	1274
5,000	366	150,000	1244
8,000	460	160,000	1212
11,000	536	180,000	1116
13,000	580	200,000	964
15,000	620	210,000	859
17,000	657	220,000	723
22,000	739	230,000	535
26,000	796	240,000	163
30,000	847		

[1] To account for sample attrition, multiply the standard error of the estimate by 1.04 for estimates which include data from Wave 5 and beyond.

Table 17. Standard Errors of Estimated Percentages of Households Families or Unrelated Persons

Base of Estimated Percentage (Thousands)	Estimated Percentages[1]					
	≤ 1 or ≥ 99	2 or 98	5 or 95	10 or 90	25 or 75	50
200	2.1	3.0	4.7	6.5	9.3	10.8
300	1.8	2.5	3.8	5.3	7.6	8.8
500	1.4	1.9	3.0	4.1	5.9	6.8
750	1.1	1.6	2.4	3.3	4.8	5.6
1,000	1.0	1.3	2.1	2.9	4.2	4.8
2,000	0.68	1.0	1.5	2.0	3.0	3.4
3,000	0.55	0.78	1.2	1.7	2.4	2.8
5,000	0.43	0.60	0.9	1.3	1.9	2.2
7,500	0.35	0.49	0.8	1.1	1.5	1.8
10,000	0.30	0.43	0.66	0.9	1.3	1.5
15,000	0.25	0.35	0.54	0.75	1.1	1.2
25,000	0.19	0.27	0.42	0.58	0.8	1.0
30,000	0.18	0.25	0.38	0.53	0.76	0.9
40,000	0.15	0.21	0.33	0.46	0.66	0.76
50,000	0.14	0.19	0.30	0.41	0.59	0.68
60,000	0.12	0.17	0.27	0.37	0.54	0.62
70,000	0.11	0.16	0.25	0.35	0.50	0.58
80,000	0.11	0.15	0.23	0.32	0.47	0.54
90,000	0.10	0.14	0.22	0.30	0.44	0.51
92,000	0.10	0.14	0.22	0.30	0.44	0.50

1 To account for sample attrition, multiply the standard error of the estimate by 1.04 for estimates which include data from Wave 5 and beyond.

Table 18. Standard Errors of Estimated Percentages of Persons

Base of Estimated Percentage (Thousands)	Estimated Percentages					
	≤ 1 or ≥ 99	2 or 98	5 or 95	10 or 90	25 or 75	50
200	3.7	5.2	8.1	11.1	16.0	18.5
300	3.0	4.2	6.6	9.1	13.1	15.1
600	2.1	3.0	4.7	6.4	9.2	10.7
1,000	1.6	2.3	3.6	5.0	7.2	8.3
2,000	1.2	1.6	2.5	3.5	5.1	5.8
5,000	0.74	1.0	1.6	2.2	3.2	3.7
8,000	0.58	0.8	1.3	1.8	2.5	2.9
11,000	0.50	0.70	1.1	1.5	2.2	2.5
13,000	0.46	0.64	1.0	1.4	2.0	2.3
17,000	0.40	0.56	0.9	1.2	1.7	2.0
22,000	0.35	0.49	0.8	1.1	1.5	1.8
26,000	0.32	0.45	0.71	1.0	1.4	1.6
30,000	0.30	0.42	0.66	0.9	1.3	1.5
50,000	0.23	0.33	0.51	0.70	1.0	1.2
80,000	0.18	0.26	0.40	0.55	0.8	0.9
100,000	0.16	0.23	0.36	0.50	0.72	0.8
130,000	0.14	0.20	0.32	0.43	0.63	0.72
200,000	0.12	0.16	0.25	0.35	0.51	0.58
220,000	0.11	0.16	0.24	0.33	0.48	0.56
230,000	0.11	0.15	0.24	0.33	0.47	0.55
240,000	0.11	0.15	0.23	0.32	0.46	0.53

1 To account for sample attrition, multiply the standard error of the estimate by 1.04 for estimates which include data from Wave 5 and beyond.

Table 19. 1991 Topical Module Generalized Variance Parameters[1]

	a	b
Fertility		
# Women	-0.0000748	6,119
Births	-0.0000670	11,158
Educational Attainment[2]		
Wave 2	-0.0000457	8,335
Wave 5	-0.0000511	9,085
Wave 8	-0.0000511	9,085
Marital Status and Person's Family Characteristics		
Some HH members	-0.0000644	12,613
All HH members	-0.0000804	15,326
Child Support		
Wave 3	-0.0000883	9,286
Support for non-household members		
Wave 3	-0.0000961	9,286
Health and Disability	-0.0000499	12,014
0-15 Child Care		
Wave 3	-0.0001340	7,514
Welfare History and AFDC		
Both sexes 18+	-0.0001241	22,040
Males 18+	-0.0002604	22,040
Females 18+	-0.0002372	22,040

1 Use the "16+ Income and Labor Force" core parameter for tabulations of reasons for not working/reservation wage and work related income.

2 The parameter also applies to the School Enrollment and Finance Topical Module Subject.

Table 20. SIPP 1990, 1991 Combined Panel Topical Module Generalized Variance Parameters

	a	b
Educational Attainment		
1990 Wave 5/1991 Wave 2	-0.0000190	3,470
1990 Wave 8/1991 Wave 5	-0.0000201	3,582
Support for non-household members		
1990 Wave 6/1991 Wave 3	-0.0000400	3,866
Health and Disability		
1990 Wave 6/1991 Wave 3	-0.0000208	5,001
0-15 Child Care		
1990 Wave 6/1991 Wave 3	-0.0000558	3,128
Child Support		
1990 Wave 6/1991 Wave 3	-0.0000368	3,866

Table 21. Distribution of Monthly Cash Income Among Persons 25 to 34 Years Old

	Total	under $300	$300 to $599	$600 to $899	$900 to $1,199	$1,200 to $1,499	$1,500 to $1,999	$2,000 to $2,499	$2,500 to $2,999	$3,000 to $3,499	$3,500 to $3,999	$4,000 to $4,999	$5,000 to $5,999	$6,000 and over
Thousands in interval	39,851	1371	1651	2259	2734	3452	6278	5799	4730	3723	2519	2619	1223	1493
Percent with at least as much as lower bound of interval	--	100.0	96.6	92.4	86.7	79.9	71.2	55.5	40.9	29.1	19.7	13.4	6.8	3.7

Table 22. SIPP Factors to be Applied to the 1991 Base Parameters to Obtain the 1990, 1991 Combined Panel Parameters[1]

Waves to be Combined

1990 panel	1991 panel	g factor[2]
5	2	0.4163
6	3	0.4163
7	4	0.4163
8	5	0.3943

[1] When deriving estimates based on two or more waves of data from the same panel, choose the corresponding g-factor with the greatest value. Apply only this factor to the base parameter.

Table 23. Factors to be Applied to Base Parameters to Obtain Combined Panel Parameters for Estimates[1] from Various Reference Periods.

# of available rotation months for 2 panels combined[2]	factor
Monthly Estimate	
2	4.0000
3	3.0000
4	2.0000
5	1.6667
6	1.3333
7	1.1667
8	1.0000
Quarterly Estimates	
12	1.8519
15	1.5631
18	1.2222
19	1.1470
24	1.0000
Annual Estimates	
96	1.0000

1 Estimates are based on monthly averages.

2 The number of available rotation months for a given estimate is the sum of the number of rotations available for each month of the estimate for the two panels. There must be at least one rotation month available for each month from each panel for monthly and quarterly estimates.